Topics

- General overview of the Fragment Molecular Orbital (FMO) method
- Examples of the types of calculations you can do with FMO
- Parallelization with GDDI
- How to fragment a system/input and output files
- Demonstration of software available

Background & Motivations

- Biomolecules contain hundreds or thousands of atoms, making accurate quantum calculations either very difficult or impossible
- QM/MM methods have become popular in recent years, however,
    - As system size grows the QM region can get unwieldy
    - The energy contribution from the environment becomes too large to obtain reasonable accuracy from molecular mechanics
- Fragmentation methods offer a unique solution to accurate calculations on large molecules

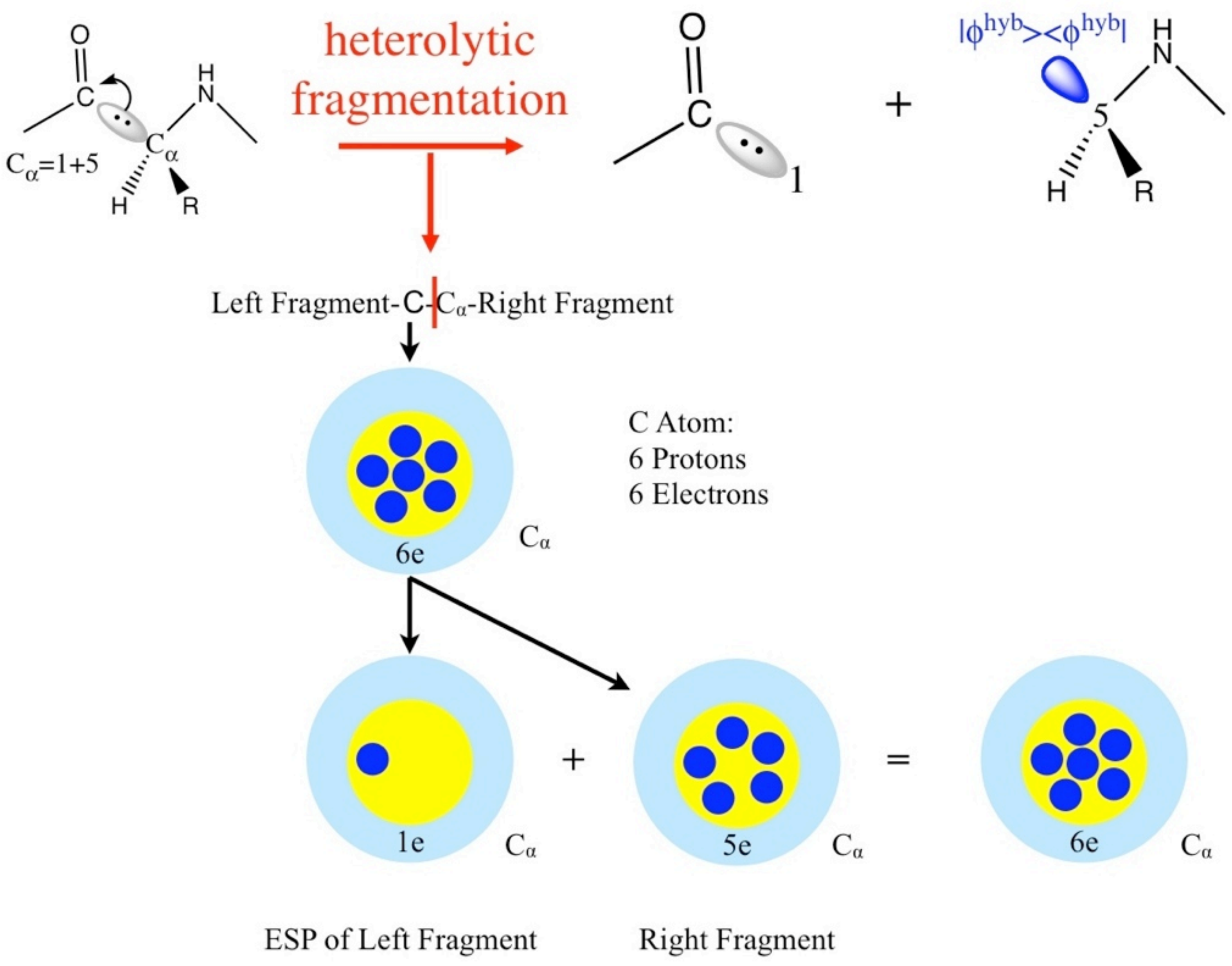
Basic Ideas

- Exchange and self-consistency are local in most molecules
    - Essentially, long range (greater than 3 angstroms) effects can be approximated by classical physics.
- We can treat long-range interactions using just the Coulomb operator (ignoring exchange)
- Perform the fragment calculations individually in the rigorous Coulomb field of the whole system
- Improved by explicit many-body corrections for pairs and triples (dimers & trimers)
- The Coulomb bath allows for fragmentation without hydrogen capping

Fragmentation Scheme

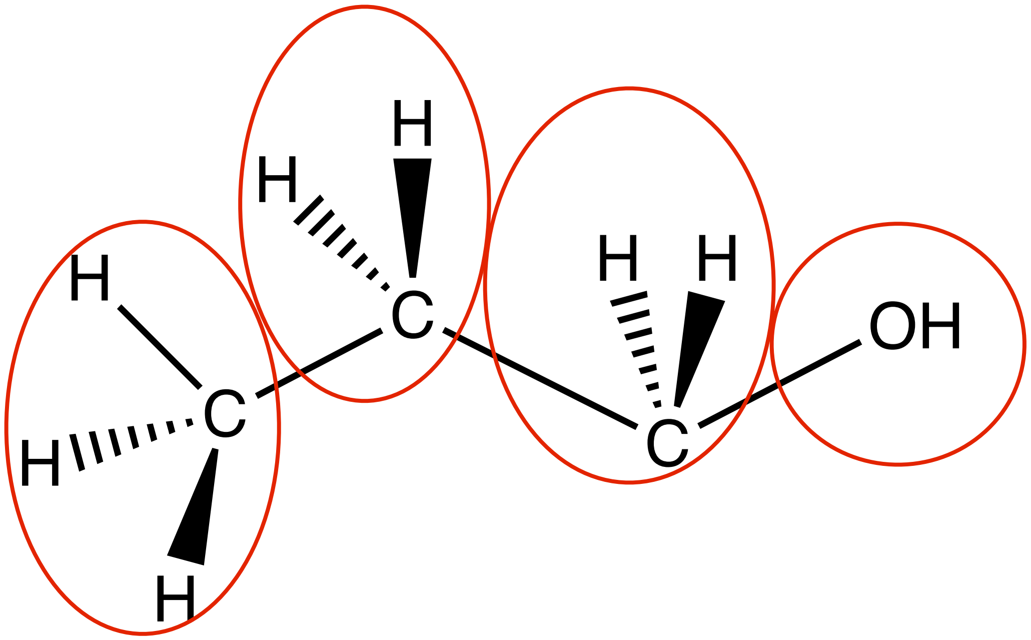
- Electrostatic fractioning of bonds
- One fragment is assigned two electrons and the other none
- FMO fragmentation is not a formal “mathematical exercise” and should be conducted based upon chemical knowledge
- Hydrogen bonding is taken into account by dimer calculations so there is no need to put pieces connected by a H-bond into the same fragment
- The dimer & trimer calculations allow for charge transfer to be taken into account

Fragmentation Scheme



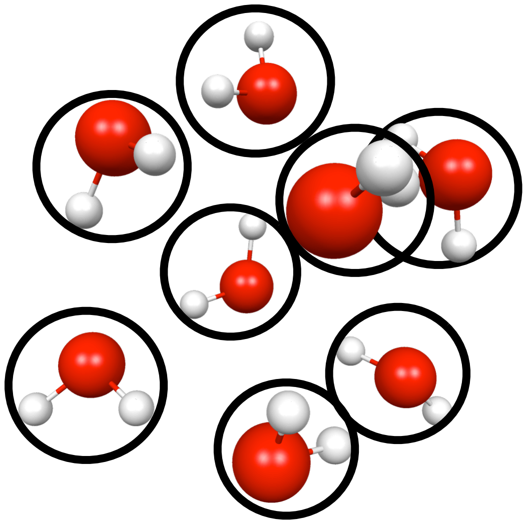
C_α=1+5
heterolytic fragmentation
$|\phi^{hyb}\rangle\langle\phi^{hyb}|$
Left Fragment-C-C_α-Right Fragment
C Atom:
6 Protons
6 Electrons
6e
C_α
1e
5e
ESP of Left Fragment
Right Fragment

FMO Fragmentation



For covalently bonded molecules, we divide the fragment into pieces so as not to destroy bond electron pairs.

In water clusters, fragmentation is easier, requiring no covalent bond breaking. We can have one water per fragment, four waters per fragment etc.



Basic FMO Methodology

1. Divide the molecule into fragments
2. Calculate the initial electron density distribution of the fragments in the Coulomb “bath” of the full system
3. Construct the individual fragment Fock operators using the densities calculated in 2 and solve for the fragment energies
4. Determine if the density has converged for all the fragments. If not, go back to step 3
5. Construct fragment pair Hamiltonians using the converged monomer densities
6. Calculate total energy and electron density

Basic FMO Methodology

The total energy of the system can be written as

$$\begin{aligned} E = {} & \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J) \\ & + \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) \\ & - (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots , \end{aligned}$$

Where the monomer (I), dimer (IJ) and trimer (IJK) energies are obtained using the standard SCF method

Examples of FMO Calculations

The picture at right is the hydrated cisplatin-DNA complex. FMO-MP2 calculations were performed using the MCP model

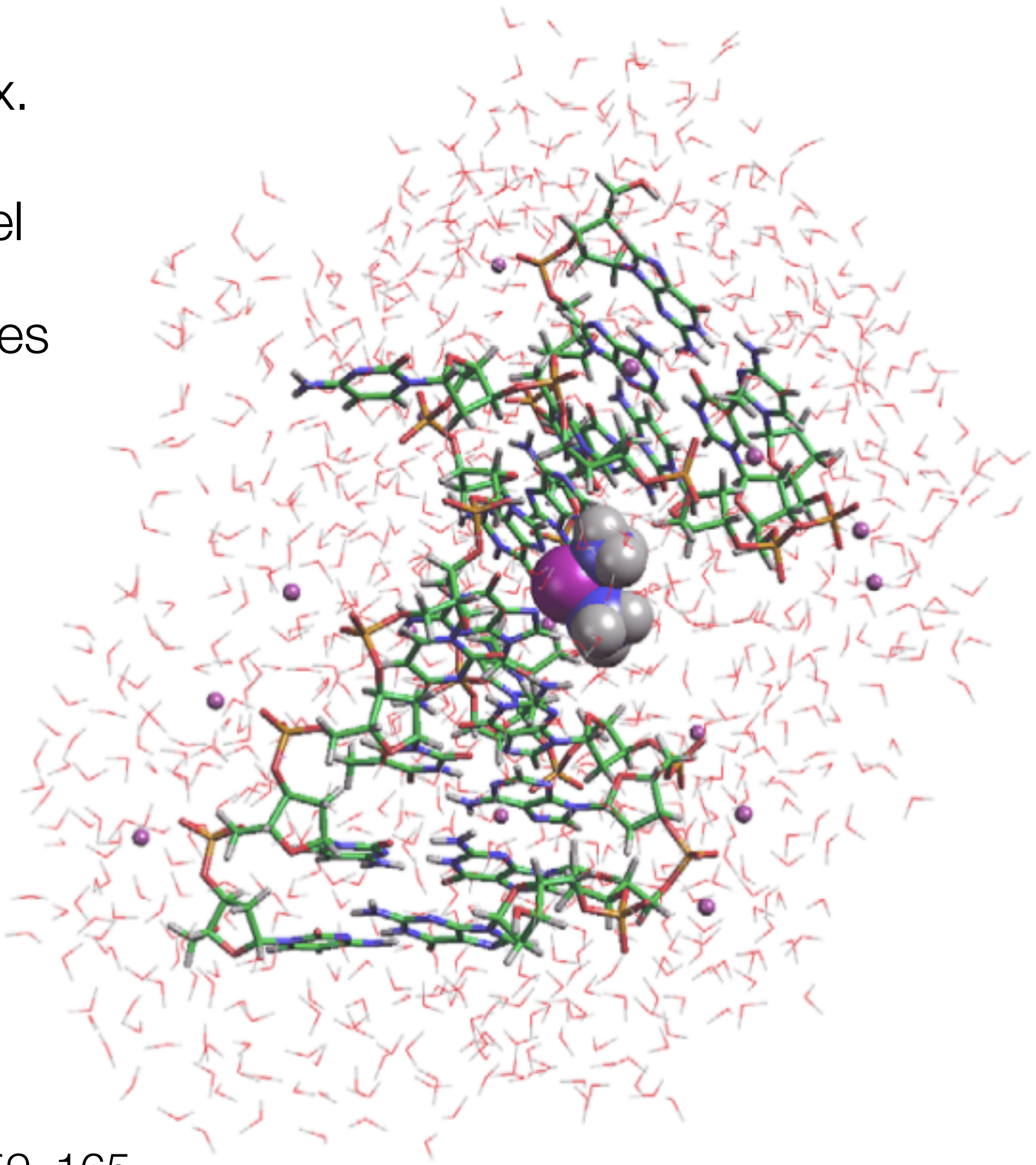
Sodium ions and water molecules were relaxed with the Amber99 force field

This calculation included:

3596 atoms

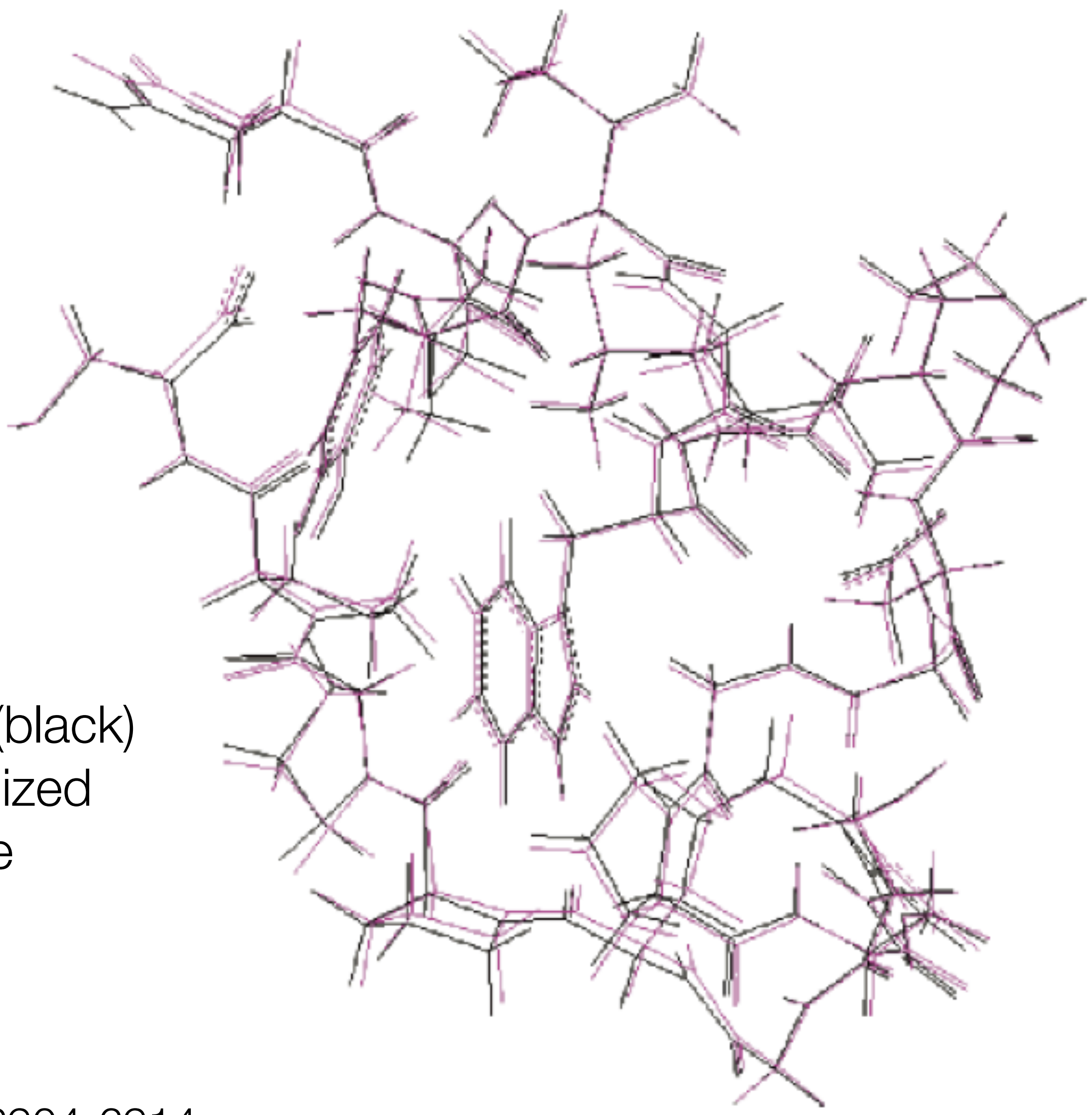
~997 fragments

10154 electrons expanded by basis functions



Chemical Physics Letters 427 (2006) 159–165

Examples of FMO Calculations



Overlay of the FMO-RHF (black) and RHF (magenta) optimized structures of the Trp-Cage protein at 3-21G

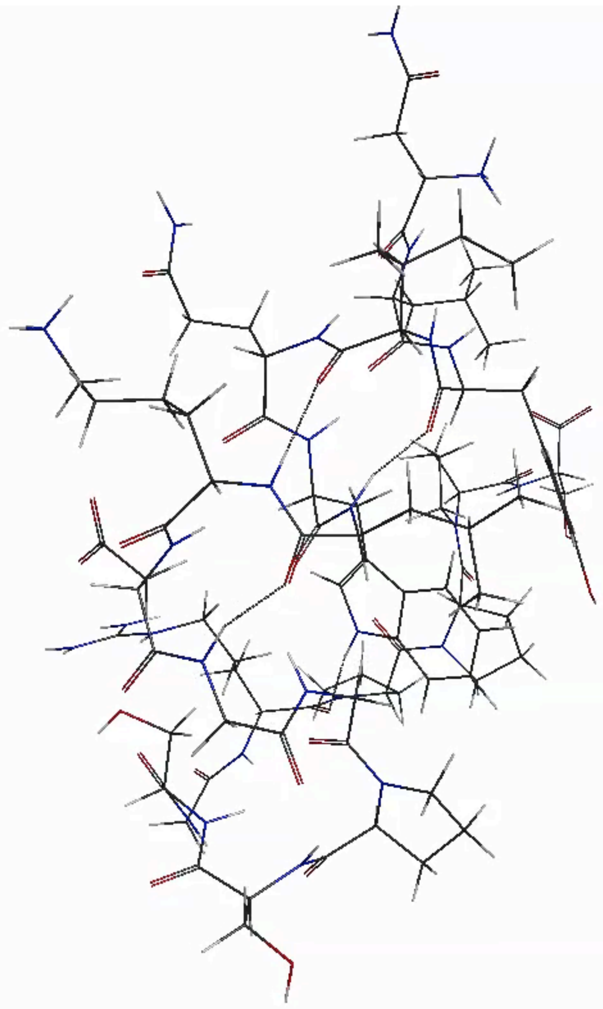
J. Phys. Chem. A 2007, 111, 6904-6914

Examples of FMO Calculations

The protein consists of 20 amino acid residues and a total of 304 atoms. The energy calculations took 160 minutes each on four nodes containing two 2.66 GHz quad core Intel Xeon processors (32 CPUs total).

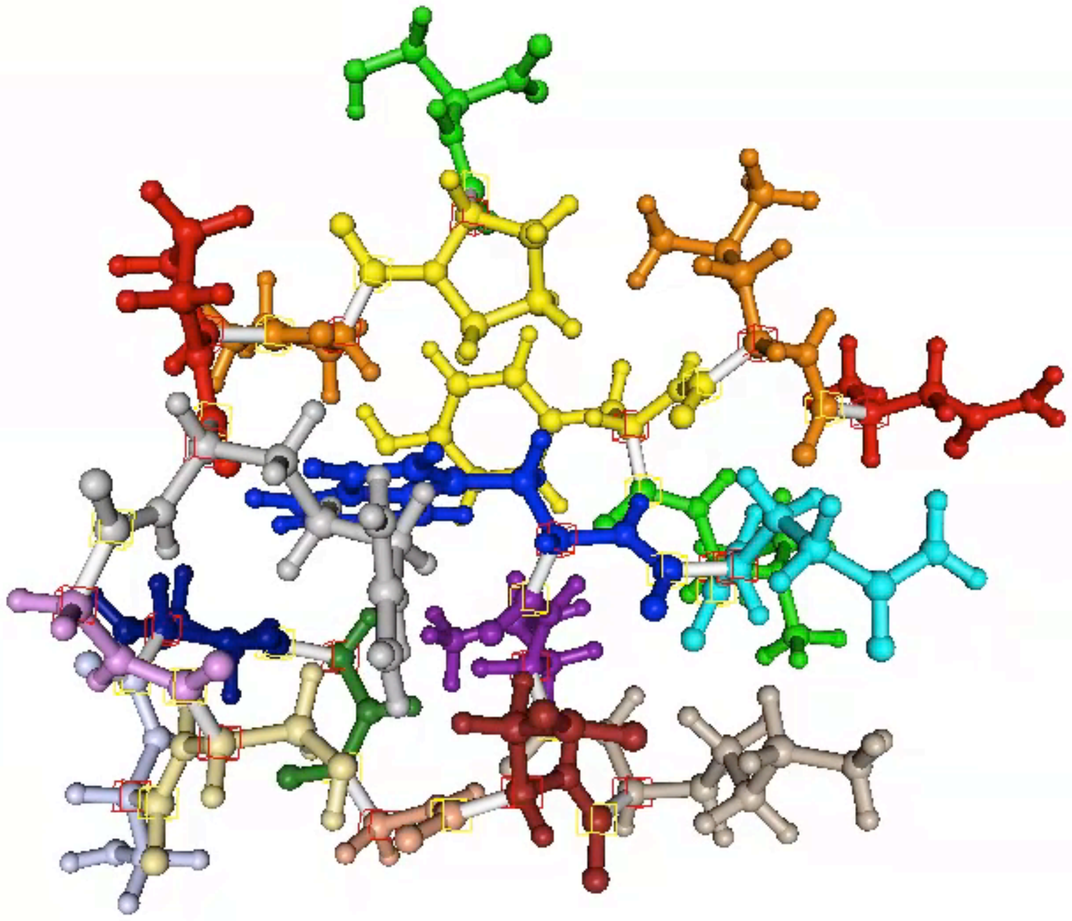
In contrast, the full ab initio ROMP2 calculation containing 2610 basis functions, if the calculation was feasible, would require 1024 CPUs with 16 GB of RAM per CPU.

The FMO2-ROMP2 excitation energy was found to be 93.35 kcal/mol (4.05 eV).



Structure of 1L2Y miniprotein

Examples of FMO Calculations



Fragmentation of of 1L2Y miniprotein

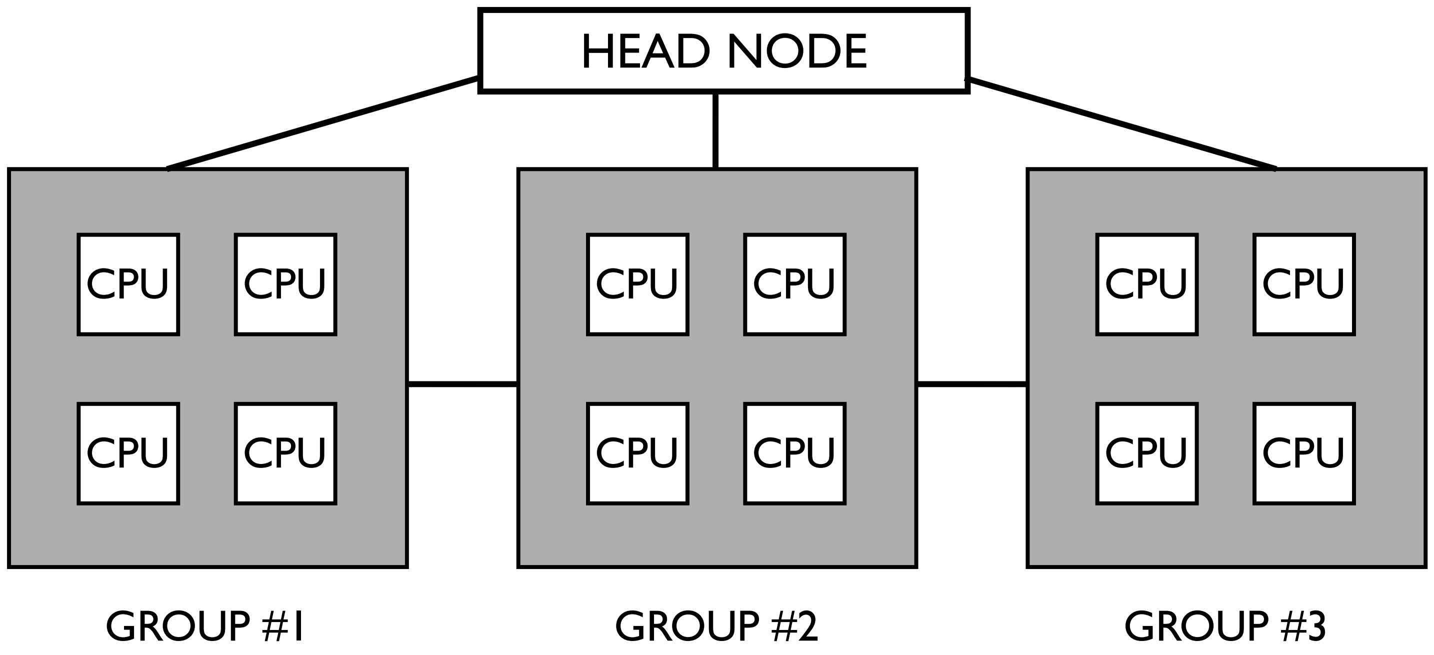
FMO is massively parallel

- As one can imagine, the fragmentation of a system into many small pieces begs to be done in parallel
- FMO was created with this in mind, and also takes advantage of the Generalized Distributed Data Interface (GDDI) implementation in GAMESS
- GDDI allows for massively parallel calculations on clusters of computers or supercomputers
    - After the molecule is divided into fragments, each fragment is sent to a group which is composed of more than one processor or SMP enclosure
    - Each fragment is then run in parallel in each group
    - This provides two levels of parallelization, greatly speeding up the calculation

J Comput Chem, 25 (2004) 872-880

The Generalized Distributed Data Interface (GDDI)

- GDDI allows for massively parallel calculations on clusters of computers or supercomputers
  - After the molecule is divided into fragments, each fragment is sent to a group which is composed of more than one processor or SMP enclosure
  - Each fragment is then run in parallel in each group
  - This provides two levels of parallelization, greatly speeding up the calculation



FMO is massively parallel

		Timing* (minutes)	
		6-31+G(d)	
<u>Tetramer</u>	#CPUs	FMO2-MP2	MP2
1-amino,4-H-1,2,4-triazolium dinitramide	8	12.2	28.4
	16	6.4	14.7
	32	3.5	7.3
<u>Hexamer</u>			
1-amino,4-H-1,2,4-triazolium dinitramide	8	24	172.1
	16	12.5	83.9
	32	6.8	42.8

* Timings for ionic liquid clusters performed on a Cray XT4 with 2.1GHz AMD Opteron64 processors.
Each node contains a 4 core CPU and 8 GB of RAM.
GDDI was not used

FMO is massively parallel

			Timing* (minutes)	
			6-31++G(d,p)	
Octamer	# Groups	#CPUs	FMO2-MP2	MP2
1-H,4-H-1,2,4-triazolium dinitramide	1	4	42.9	-
	1	8	22.3	-
	2	16	11.4	-
	4	32	5.8	-
	8	64	3.5	55.6

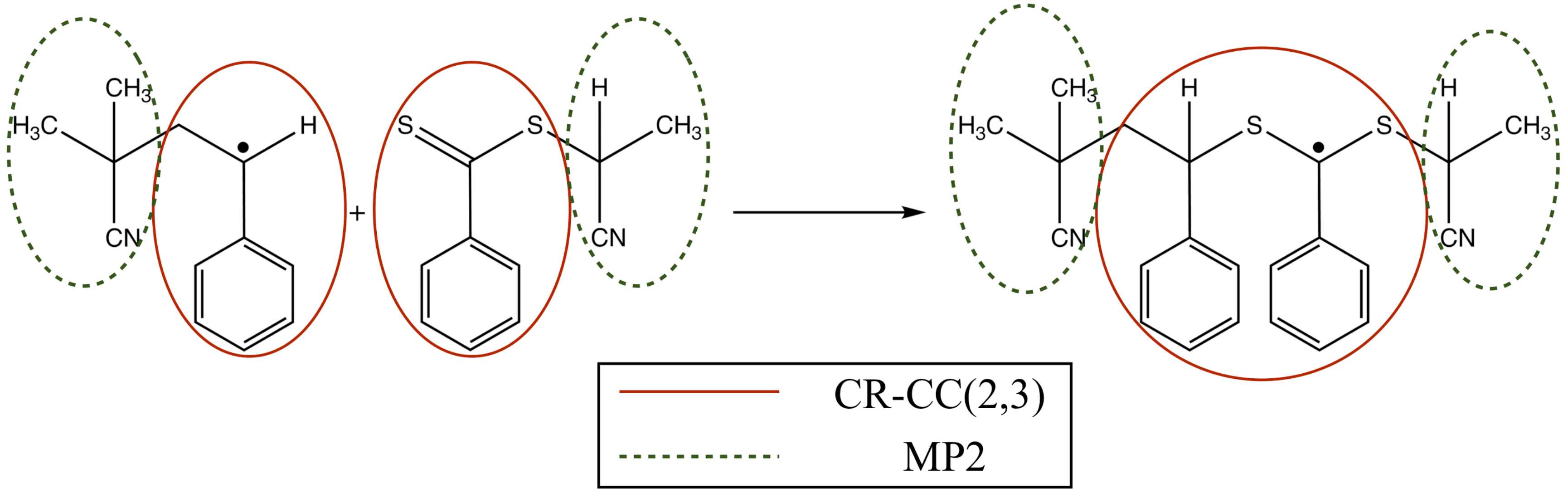
Full ab initio MP2 calculation requires ~77 GB of RAM
FMO2-MP2 calculation requires less than 1 GB of RAM
Error = 0.67 kcal/mol

* Timings for ionic liquid clusters performed on a computer cluster containing 2.66GHz Intel Xeon processors.
Each node contains two 4 core CPUs and 16 GB of RAM.
GDDI division was across individual nodes (1 group = 8 CPUs)

FMO has many other methods implemented already

- FMO has all of the following wavefunction types implemented
    - RHF, ROHF, DFT, MP2, CC and MCSCF
      (all of which except MCSCF support FMO3)
- FMO also has a multilayer implementation allowing you to specify difference levels of electron correlation or basis sets within different layers
- FMO is also interfaced with PCM and EFP for solvent effects
- RUNTYPs of interest are ENERGY, GRADIENT, OPTIMIZE, OPTFMO

Example of a Multi-Layer FMO (MFMO) Calculation



Fragmentation Options

- If you like proteins, then you can look in the PDB database and use “fmoutil” to do the fragmentation for you
  - Typical fragmentation for a protein is one or two residues per fragment
- If you like pure clusters of molecules, simply use NACUT= in $FMO to tell the program how many atoms per fragment
  - e.g. water clusters would have NACUT=3, assuming you have the coordinates in a sensible order

Other Fragmentation Options

- You can use the free (Windows only) program Facio for automated protein fragmentation
- The process of fragmentation with Facio is fairly easy
- Detailed instructions for how to create an FMO input with Facio can be found at: *http://staff.aist.go.jp/d.g.fedorov/fmo/facio-fmo.html*
- If you want to do something other than a protein or pure cluster of molecules, you're on your own (i.e. use your chemical intuition)

Other Fragmentation Options

- Instructions can be found in the GAMESS directory (gamess/tools/fmo/tutorials)
- The file basic.inp is quite helpful in understanding the way to fragment “by hand”
- There is also a file multilayer.inp which describes how to perform multilayer FMO calculations

FMO Input

$FMO group contains FMO parameters:

- NFRAG = # fragments
- NBODY = 2 or 3
- NACUT = Number of Atoms to Cut for each fragment
- RESPAP, RESPPC, RESDIM, RCORSD, RITRIM, MODESP all control how far separated fragments and the electrostatic potential
- SCFTYP = specifies RHF, ROHF etc for each fragment
- ICHARG = specifies fragment charges
- MULT = specifies fragment multiplicities

FMO Input

- $FMOPRP group contains options for printing (NPRINT) and GDDI group division (NGRFMO)
- $FMOHYB contains the hybrid orbitals used for bond division. Orbitals for common basis sets can be found in the gamess/tools/fmo/HMO/HMO.txt file
- $FMOXYZ contains the cartesian coordinates for the system

FMO Input

Demonstration of different fragmentation tools:

- Facio
- NACUT scheme
- Manual input creation

FMO Output

Energies at the end of the .log file look like (for FMO3):

```
Total energy of the molecule: Ecorr  (3)=     -458.197699251

 Total energy of the molecule: Euncorr(3)=     -456.438421958

 Total energy of the molecule: Edelta (3)=       -1.759277293
```

FMO Problems

Systems that have poor accuracy:

- Metallic crystals
- Large metallic clusters
- Single molecule fullerenes
- Pretty much anything that has delocalized electrons

However, if the system is large enough, the area with delocalized electrons can be placed in a fragment by itself